LEVEL II

Research Memorandum 72-7

SOME FACTORS AFFECTING MENSURATION VARIABILITY AMONG IMAGE INTERPRETERS

ADA 079387

BESRL

U. S. Army
Behavior and Systems Research Laboratory

September 1972

Army Project Number
(16) 2Q662704A732

Image Interpretation Displays c-14

(14) BESRL-RM-

(9) Research ~~Memorandum~~ 72-7

(6) SOME FACTORS AFFECTING MENSURATION VARIABILITY AMONG IMAGE INTERPRETERS.

(10) J. Richard Lepkowski
System ~~Development~~ Corporation
and Thomas E. Jeffrey BESRL

Abraham H. Birnbaum, Program Director

A

Submitted by:
Joseph Zeidner, Chief
Support Systems
Research Division

(12) 19

Approved by:
J. E. Uhlaner, Director
U. S. Army Behavior and
Systems Research Laboratory

(11) Sept~~ember~~ 1972

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the Behavior and Systems Research Laboratory.

402 777 LB

SOME FACTORS AFFECTING MENSURATION VARIABILITY AMONG IMAGE INTERPRETERS

BACKGROUND

Measurement of object dimensions and distance is one of the most important tasks in image interpretation. The image interpreter determines imagery scale through the measurement of objects of known length and/or width; computes volume using measurements of the length, breadth, and height of objects; triangulates by measurement of distance between points; uses measurement to aid in the identification of unknown objects, and makes many other determinations by mensuration. These measurements must be made rapidly and accurately. To this end, equipment has been designed and provided to assist him.

Among the aids provided for image interpretation are graduated scales (in several different units of measure) and tube magnifiers (with a range of magnifications). A preliminary investigation revealed that repeated measurement of the same object using different scales and magnifications produced a range of values. That such differences in the measurement of the same object is encountered in practice is attested to by the fact that the average value of measurements made by two or more image interpreters is frequently used as the best estimate of the true image size of an object.

OBJECTIVES

Variability in object measurement could be attributed to individual techniques of image interpreters, differences in choice of measuring scale, magnification power, object size, imagery scale, or some combination of these. The specific objectives of this exploratory study were:

(1) To establish the mensuration performance of experienced image interpreters.

(2) To determine which combination of scale and/or magnification power results in the least amount of mensuration variability.

(3) To determine how imagery scale and/or object size influence mensuration variability.

METHOD

EQUIPMENT

The equipment used in this study consisted of a standard light table and one each of the available scales and tube magnifiers normally used in mensuration. Specifically, these included:

1. Interpreter scale - A transparent plastic strip with two different scales etched thereon--a .001 foot interval scale and a .5 millimeter interval scale.

2. Tube magnifiers of three different powers--2-power, 7-power, and 12-power.

3. Two 7-power tube magnifiers--one with the reticle graduated in .001 foot increments and the other with the reticle graduated in .1 millimeter increments.

IMAGERY

Two frames of annotated imagery were selected at each of three different scale levels--approximately 1:2,000, 1:5,000, and 1:10,000. At each scale level, the two exposures included one exposure containing a small target (¼-ton truck or small tower base) and the other containing a large target (2½-ton truck). Each target was selected such that the sun was perpendicular, or nearly so, to the axis of the longest dimension of the target. Thus, the extremities of the targets were clearly defined.

SUBJECTS

Seven BESRL image interpreters with 10 or more years of experience performed the mensuration tasks in this experiment.

TASK

Each image interpreter was required to measure all 6 annotated targets presented one to a frame in keeping with the following schedule:

Frame	Target Size	Scale
1	small	large
2	large	large
3	small	intermediate
4	large	intermediate
5	small	small
6	large	small

A total of 4 runs through the imagery was made by each interpreter. In the first run, each target in turn was measured using the .001 foot scale with the 2-power tube magnifier followed by measurement with the .5mm scale with the 2-power tube magnifier. Run two was made using a 7-power tube magnifier with the two scales following the order given above. The third run utilized a 12-power tube magnifier with the two scales. The fourth, and final run, employed a 7-power tube magnifier with a .001 foot reticle first and then a 7-power tube magnifier with a .1mm reticle for the measurement of each target.

All measurements were recorded on forms provided the interpreters but reference back to previously recorded measurements was not permitted. Since measurement precision was stressed in this experiment, the interpreters were allowed as much time as they desired for making their measurements. No record was made of the time required for each measurement.

INDEPENDENT VARIABLES

Four independent variables were manipulated in this experiment. These were:

1. Target size - 2 levels (small--¼ ton truck or small tower base
   (large--2½ ton truck

2. Interpreter Scale - 2 levels (with .001 foot graduations
   (with .5mm graduations

3. Magnification - 3 levels (2-power
   (7-power
   (12-power

4. Imagery scale - 3 levels (about 1:2,000
   (about 1:5,000
   (about 1:10,000

DEPENDENT VARIABLE

The primary purpose of this research was to determine some sources of variability in mensuration performance. The standard deviation was used as the measure of variability:

$$\text{Standard Deviation (SD)} = \frac{\left[N \Sigma X^2 - (\Sigma X)^2 \right]^{\frac{1}{2}}}{N}$$

ANALYSIS

Two separate analyses were performed. The first involved all four independent variables; the second, which used data obtained using the 7-power tube magnifiers equipped with two different reticles, involved only three independent variables since magnification was a fixed value.

In order to carry out the desired analyses of variability, it was necessary to have all measurements expressed in the same unit of measure. This was accomplished by converting all of the .001 foot measurements to their equivalent millimeter values. Appendix A lists the original values of the measurements made by the seven subjects for each of the treatment conditions. Appendix B gives similar data for the three-factor experiment.

The analysis of variance tests computed in this exploratory research effort do not provide a proper error estimate. Therefore, the interactions having small mean square values were combined as an estimate of the error. The sum of squares for the selected interactions were summed and divided by the combined degrees of freedom to obtain the mean square for the error term. This procedure depends upon the implicit assumption that the pooled interactions are not significant. If this assumption is valid, the pooled interactions provide a reasonable estimate of the experimental error term. As previously stated, what is being analyzed is not the individual measurements made by the 7 subjects but the standard deviation of these measurements for each of the various treatment conditions.

RESULTS

FOUR-FACTOR ANALYSIS

Table 1 shows the 36 treatment means for the four-factor analysis and Table 2 gives the standard deviations showing the interpreter variability about these means. These standard deviations and means were determined from the measurements made by all 7 image interpreters under each of the 36 experimental conditions with one exception. For the condition in which the interpreters measured a large target on imagery of the largest scale using an interpreter scale graduated in thousandths of a foot and 2-power

Table 1

TREATMENT MEANS FOR FOUR-FACTOR ANALYSIS
(in millimeters)

Magnification Level	Interpreter Scale Graduation	Small Target			Large Target		
		S_1	S_2	S_3	S_1	S_2	S_3
2-power	.001 foot	1.833	.618	.300	3.937	1.398	.853
	.5 mm	1.893	.621	.333	3.986	1.361	.786
7-power	.001 foot	1.898	.605	.370	3.941	1.433	.884
	.5 mm	1.864	.643	.377	4.000	1.437	.864
12-power	.001 foot	1.859	.618	.357	3.932	1.446	.853
	.5 mm	1.886	.650	.389	3.997	1.457	.831

Table 2

TREATMENT STANDARD DEVIATIONS FOR FOUR-FACTOR ANALYSIS
(in millimeters)

Magnification Level	Interpreter Scale Graduation	Small Target			Large Target		
		S_1	S_2	S_3	S_1	S_2	S_3
2-power	.001 foot	.0771	.0137	.0253	.0567	.0969	.0710
	.5 mm	.0776	.0589	.0552	.0350	.0751	.0350
7-power	.001 foot	.1015	.0253	.0552	.0917	.0863	.0326
	.5 mm	.0789	.0416	.0266	.0535	.0462	.0789
12-power	.001 foot	.0488	.0268	.0482	.0430	.0585	.0488
	.5 mm	.0789	.0655	.0554	.0539	.0678	.0514

magnification, the measurement reported by the second interpreter in the list given in Appendix A was judged "unbelievable." The measurement reported was .0150 feet almost .002 feet greater than that reported by the other six interpreters. This is two graduations on the interpreter scale. It was judged that this departure from the average of the group was due to some temporary aberration of this interpreter that was not related to the measuring instruments used or the procedures employed by the interpreters in making these measurements. Consequently, this one score was not used and the mean and standard deviation for this one experimental condition was determined on the basis of six interpreters rather than seven.

Table 3 gives the summary for the analysis of variance of the standard deviations of the obtained measurements for the four-factor experiment--interpreter scale, magnification, target size, and imagery scale. None of the main effects produced a significant difference in performance variability. The mean standard deviations for each level of the independent variables are given in Table 4. Note the small absolute differences among these means.

The target size by imagery scale (T x S) interaction produces the only statistically significant effect on performance variability for this analysis. Table 5 gives the mean standard deviations for this interaction. An examination of the means does not reveal any pattern that suggests an explanation for this significant result. Figure 1 was plotted to obtain

Table 3

SUMMARY TABLE FOR ANALYSIS OF VARIANCE OF STANDARD DEVIATIONS

Source of Variance	Sum of Squares(10^6)	df	Mean Square(10^6)	F	$F_{.95}$	$F_{.99}$
R (Interpreter Scale)	21.7778	1	21.7778	.0668	4.22	7.72
M (Magnification)	213.2939	2	106.6470	.3271	3.37	5.53
T (Target Size)	412.0900	1	412.0900	1.2640	4.22	7.72
S (Imagery Scale)	1932.3889	2	966.1944	2.9636	3.37	5.53
R x M	1165.9105	2	582.9552	------		
R x T	1171.9211	1	1171.9211	3.5946	4.22	7.72
R x S	345.2822	2	172.6411	------		
M x T	208.2150	2	104.1075	------		
M x S	1494.4011	4	373.6003	------		
T x S	4498.3467	2	2249.1734	6.8989**	3.37	5.53
R x M x T	1069.3006	2	534.6503	------		
R x M x S	889.9345	4	222.4836	------		
R x T x S	1046.0955	2	523.0478	------		
M x T x S	641.4234	4	160.3558	------		
R x M x T x S	1615.9677	4	403.9919	------		
TOTAL	16726.3489	35				
POOLED ERROR TERM[a]	8476.5305	26	326.0204			

** Means differ significantly, $P < .01$.

[a] Pooled error term obtained by summing values of sum of squares for those sources indicated by (------) in F column.

a pictorial representation of the interaction effects. The actual targets measured by the image interpreters were examined and a subjective estimate made of the relative sharpness of the edge definition of each target. These estimates appear in parentheses by each point plotted in Figure 1. It appears possible that image sharpness may be confounded with the independent variables of this experiment. Although the edge definition estimates are subjective, they suggest that mean change in performance may have been due to variations in image quality as well as, or instead of, changes in imagery scale and/or target size.

Table 4

MEAN STANDARD DEVIATION FOR EACH MAIN EFFECT LEVEL

Variable	Level	Mean (mm)
Interpreter Scale	.001 foot graduations	.056
	.5 millimeter graduations	.058
Magnification	2-power	.056
	7-power	.060
	12-power	.054
Target Size	small target	.053
	large target	.060
Imagery Scale	1: 2,000	.066
	1: 5,000	.055
	1:10,000	.049

Table 5

STANDARD DEVIATION CELL MEANS FOR TARGET SIZE X IMAGERY SCALE INTERACTION

Target Size	Imagery Scale		
	1:2,000	1:5,000	1:10,000
Small	.077	.039	.044
Large	.056	.072	.053

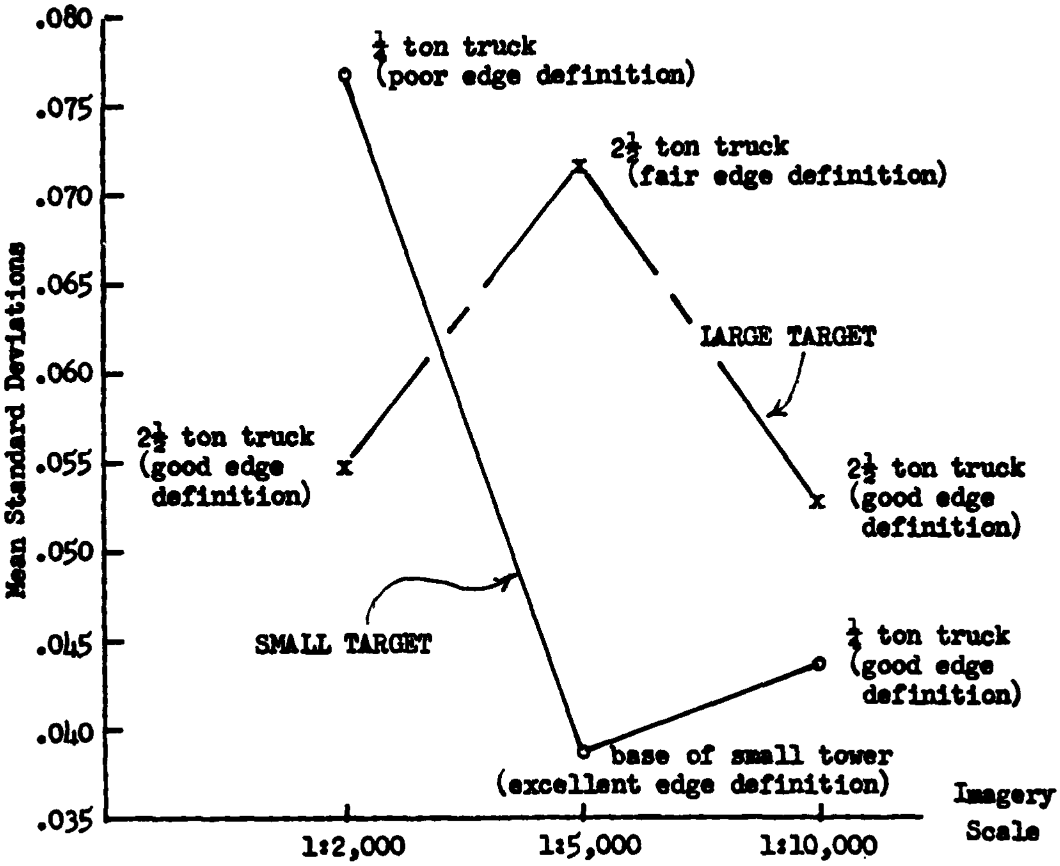

Figure 1. Sketch of Target Size by Imagery Scale Interaction

THREE-FACTOR ANALYSIS

A separate experiment was conducted using the same imagery, subjects, and targets but requiring that measurements be made by means of two reticles--measurements were made using reticle graduated in thousandths of a foot first followed by measurements made with reticle graduated in tenths of a millimeter--each reticle attached to a 7-power tube magnifier.

Table 7 summarizes the analysis of variance of the standard deviations of the measurements made under the several experimental conditions. No conventional error term was available since there were no replications. The second-order interaction term was used as the best estimate of the error term and it was found that none of the interactions approached significance. Therefore, all interactions were grouped into a single estimate of the error and the pooled sum of squares divided by 7 degrees of

Table 6

TREATMENT MEANS AND STANDARD DEVIATIONS FOR THREE-FACTOR ANALYSIS (in millimeters)

Statistic	Reticle Graduation	Small Target			Large Target		
		S_1	S_2	S_3	S_1	S_2	S_3
Mean	.001 foot	1.859	.618	.318	3.928	1.398	.827
	.1 mm	1.850	.610	.343	3.913	1.413	.849
Standard Deviation	.001 foot	.0564	.0138	.0320	.0254	.0444	.047[illegible]
	.1 mm	.0393	.0521	.0291	.0365	.0328	.03[illegible]

freedom. None of the main effects has a significant effect on the variability of subject measurement performance. This statement assumes that the pooled interactions provide a valid estimate of the experimental error in this experiment.

Table 7

SUMMARY TABLE FOR ANALYSIS OF VARIANCE OF STANDARD DEVIATIONS

Source of Variance	Sum of Squares(10^6)	df	Mean Square(10^6)	F	$F_{.95}$	$F_{.99}$
R (Reticle)	6.0208	1	6.0208	.0278	5.59	12.25
T (Target Size)	.2408	1	.2408	.0011	5.59	12.25
S (Imagery Scale)	28.8817	2	14.4408	.0668	4.74	9.55
R x T	65.8008	1	65.8008	-----		
R x S	218.4117	2	109.2058	-----		
T x S	463.7017	2	231.8508	-----		
R x T x S	765.7517	2	382.8758	-----		
TOTAL	1548.8092	11				
POOLED ERROR TERM[a]	1513.6658	7	216.2380			

[a] Pooled error term obtained by summing values for sum of squares for sources indicated by (-----) in F column.

INTRA-INTERPRETER MENSURATION CONSISTENCY

To estimate the consistency with which the interpreters provided high or low measurements for the same target regardless of the measurement tools employed, intercorrelations between selected measurement conditions were determined. The first of these analyses correlated the measurements made by the 7 interpreters using an interpreter scale with .001 foot graduations and a 7-power tube magnifier with the measurements made by the same interpreters using a 7-power tube magnifier equipped with a reticle graduated in thousandths of a foot. A second analysis correlated measurements made with an interpreter scale graduated in .5 millimeters using a 7-power tube magnifier with measurements made using a reticle with .1 millimeter graduations on a 7-power tube magnifier. The final analysis correlated measurements made using a reticle graduated in .001 foot intervals on a 7-power tube magnifier with measurements made using a reticle graduated in .1 millimeter intervals on a 7-power tube magnifier. Table 8 shows the rank difference correlations obtained.

Table 8

RANK DIFFERENCE CORRELATIONS AMONG REPEATED MEASUREMENTS OF 7 INTERPRETERS

Measurements with 7-power Tube Magnifier and	Imagery Scale		
	S_1 (1:2,000)	S_2 (1:5,000)	S_3 (1:10,000)
<u>.001' Scale vs .001' Reticle</u>			
Small Target	.46	.74	.44
Large Target	.57	.34	.43
<u>.5mm Scale vs .1mm Reticle</u>			
Small Target	.42	.53	.59
Large Target	.54	.22	.42
<u>.001' Reticle vs .1mm Reticle</u>			
Small Target	.77	.67	.61
Large Target	.52	.33	.63

These rank difference correlations are based on 7 subjects. For a situation in which two variables are uncorrelated in the population but ranks are assigned to observations made on these variables for a limited set of observations, the correlations that might be obtained by chance have been determined empirically for sets as large as eight. For a single rank difference correlation determined from seven observations on each variable to reach the 5 percent level of confidence, the coefficient must be .75 or larger. Table 8 shows only one coefficient that reaches this level. However, the 18 correlations computed are all positive and are all moderately large. This degree of consistency among the magnitudes of the rank difference correlation coefficients leads to the conclusion that there is a tendency among these seven image interpreters to maintain the same relative position in their measurement performance for the various targets and experimental conditions. The same interpreter tends to come up with the greatest value, the intermediate value, the smallest value, or some relatively consistent value regardless of the target he is measuring or the tools he is using. Techniques and procedures used by the individual interpreter may be responsible for this tendency.

INTERPRETER SCALE VERSUS RETICLE

Does measurement variability differ as a function of the mensuration tool employed? This question is not answered by the two variance analyses reported. The four-factor analysis employed the interpreter scale as the mensuration tool while the three-factor analysis used the tube magnifier reticle for measurement purposes. The degree of measurement variability is not the only consideration in determining the relative merit of these devices. They have unique applications as well as general areas of usefulness.

The 7-power tube magnifier fitted with the .001 foot reticle can be used to measure the extent of an object that does not exceed .060 feet in a single measurement. If the object dimension exceeds this value, its length must be determined in a piecemeal fashion and the size of the increments added. Such a procedure will contribute to the measurement error since the marking of the extremities of each increment will increase the sources for error.

The interpreter scale graduated in thousandths of a foot can be used to determine the length of objects that exceed a foot in extent--several versions of this scale have been issued but one version reads to 1.1 feet. Suppose that an airfield runway is recorded on a piece of imagery and the length of this runway is desired. It will probably be quicker and more accurate to measure the length of the runway using the interpreter scale than by using the 7-power tube magnifier with reticle. If the length of the runway on the imagery were 1.1 feet it could be measured in one step using the interpreter scale while 19 discrete increments would have to be measured if the reticle were used. Obviously, the magnitude of the target will be a determining factor in deciding which is the better scale to use. For the targets measured in the experiments reported here, the capacity of the measurement tool never became a limiting factor.

Table 9 repeats measurement variability data for the two mensuration tools discussed. These data are taken from two tables appearing in previous sections of this report. From Table 2 the standard deviations of measurements made using the interpreter scale with .001 foot graduations and a 7-power tube magnifier are repeated while from Table 6 the standard deviations of measurements made using the 7-power tube magnifier with .001 foot reticle are given. No statistical test for the significance of the tabled values was made. It can be seen that the variability appears to be less when a reticle is used than when the interpreter scale is used. The mean across imagery scale and target size is almost half as large when the reticle is used than it is when the interpreter scale is used.

Table 9

COMPARATIVE MEASUREMENT VARIABILITY FOR INTERPRETER SCALE AND RETICLE

7-Power Tube Magnifier &	Small Target			Large Target			Mean
	S_1	S_2	S_3	S_1	S_2	S_3	
Interpreter Scale .001 foot graduations	.1015	.0253	.0552	.0917	.0863	.0326	.0654
Reticle .001 foot graduations	.0564	.0138	.0320	.0254	.0444	.0473	.0366

CONCLUSIONS

Within the constraints imposed by the experiments conducted--imagery scale, target size, mensuration task, and measurement tools--the following conclusions appear to be justified:

1. The use of interpreter scales or reticles graduated in thousandths of a foot or in millimeters has no significant effect on mensuration variability. The foregoing should not be construed to indicate that both are unnecessary. Until some final decision is made and implemented concerning the adoption of an international unit of measurement, maps, interpretation keys, and users will require that measurements be made in the metric system and/or the English-speaking system. However, one may eliminate graduations within the same measurement system, i. e., .5mm and .1mm in the metric system or .001 foot and .0005 foot in the other system.

2. Measurement variability for targets of the size used in these experiments does not vary significantly with target ground size.

3. Imagery scale has no significant influence on the variability of target measurement. The significant interaction obtained in the four-factor analysis between target size and imagery scale was probably an artifact.

4. Magnification level has no significant effect on mensuration variability.

5. Measurements made using reticles appear to be less variable than those made with an interpreter scale. This difference was not tested statistically.

6. Interpreters tend to maintain their relative position from measurement task to measurement task with respect to the mean measurement of the group--the individual may be above average, average, or below average. This may be the result of individually acquired mensuration techniques.

APPENDIX A

INTERPRETER MEASUREMENTS FOR FOUR-FACTOR EXPERIMENT

The individual measurements of 7 subjects using the interpreter scale with .001 foot and .5 millimeter graduations for small and large size targets at large (S_1), intermediate (S_2), and small (S_3) imagery scales using 2-power, 7-power, and 12-power magnification.

.001' Ruler

2X

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
.0060	.0020	.0010	.0130	.0045	.0030
.0064	.0021	.0010	.0150	.0050	.0024
.0055	.0020	.0010	.0130	.0050	.0030
.0060	.0020	.0010	.0130	.0045	.0025
.0062	.0020	.0011	.0125	.0046	.0028
.0060	.0021	.0008	.0130	.0040	.0030
.0060	.0020	.0010	.0130	.0045	.0029

7X

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
.0060	.0020	.0010	.0130	.0045	.0030
.0061	.0021	.0013	.0131	.0051	.0030
.0068	.0020	.0015	.0134	.0050	.0030
.0060	.0020	.0010	.0130	.0047	.0029
.0067	.0020	.0014	.0125	.0048	.0029
.0060	.0020	.0011	.0125	.0042	.0028
.0060	.0018	.0012	.0130	.0046	.0027

APPENDIX A (Continued)

12X

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
.0060	.0020	.0010	.0130	.0045	.0025
.0062	.0022	.0012	.0126	.0051	.0030
.0062	.0021	.0015	.0130	.0045	.0030
.0060	.0020	.0011	.0128	.0047	.0028
.0064	.0020	.0012	.0129	.0048	.0028
.0059	.0020	.0010	.0130	.0048	.0028
.0060	.0019	.0012	.0130	.0048	.0027

.5mm Ruler

2X

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
1.90	.65	.40	4.00	1.33	.70
1.80	.70	.30	4.00	1.30	.80
2.00	.50	.25	4.00	1.40	.80
1.85	.65	.30	4.00	1.35	.80
1.90	.60	.40	4.00	1.40	.80
1.80	.60	.30	3.90	1.25	.80
2.00	.65	.38	4.00	1.50	.80

APPENDIX A (Continued)

7X

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
1.90	.60	.40	3.90	1.40	.80
1.80	.70	.40	4.10	1.51	1.00
2.00	.60	.40	4.00	1.50	.90
1.80	.65	.34	4.00	1.40	.80
1.90	.70	.40	4.00	1.45	.95
1.75	.65	.35	4.00	1.40	.80
1.90	.60	.35	4.00	1.40	.80

12X

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
1.90	.60	.40	3.90	1.40	.75
2.00	.75	.47	4.00	1.60	.85
1.90	.60	.40	4.00	1.50	.90
1.80	.60	.30	3.98	1.40	.82
1.95	.75	.45	4.00	1.45	.80
1.75	.60	.35	4.10	1.45	.90
1.90	.65	.35	4.00	1.40	.80

APPENDIX B

INTERPRETER MEASUREMENTS FOR THREE-FACTOR EXPERIMENT

The individual measurements of 7 subjects using a 7-power tube magnifier reticle with .001 foot and .1 millimeter graduations for small and large size targets at large (S_1), intermediate (S_2), and small (S_3) imagery scales.

.001' Reticle

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
.0060	.0020	.0010	.0129	.0045	.0027
.0060	.0021	.0013	.0129	.0048	.0029
.0062	.0020	.0010	.0130	.0045	.0029
.0060	.0020	.0010	.0128	.0044	.0027
.0061	.0021	.0010	.0128	.0046	.0025
.0059	.0020	.0010	.0128	.0048	.0025
.0065	.0020	.0010	.0130	.0045	.0028

.1mm Reticle

Small Target			Large Target		
S_1	S_2	S_3	S_1	S_2	S_3
1.80	.60	.30	3.90	1.40	.80
1.87	.68	.37	3.91	1.49	.90
1.85	.50	.38	4.00	1.40	.90
1.83	.62	.30	3.90	1.41	.85
1.90	.62	.35	3.88	1.38	.80
1.80	.60	.35	3.90	1.41	.85
1.90	.65	.35	3.90	1.40	.84